Ecological Informatics 78 (2023) 102332

Contents lists available at ScienceDirect

Ecological Informatics

journal homepage: www.elsevier.com/locate/ecolinf

Mining crowdsourced text to capture hikers' perceptions associated with landscape features and outdoor physical activities

Abdesslam Chai-allah [a,*], Nathan Fox [b], Fritz Günther [c], Fadila Bentayeb [d], Gilles Brunschwig [a], Sandro Bimonte [e], Frédéric Joly [a]

[a] University of Clermont Auvergne, INRAE, VetAgro Sup, UMR Herbivores, 63122 Saint-Genès-Champanelle, France
[b] University of Michigan, School for Environment and Sustainability, Ann Arbor, MI 48109, USA
[c] Department of Psychology, Humboldt Universität zu Berlin, 10099 Berlin, Germany
[d] University of Lyon, Lyon2, ERIC, EA 3083, 5 avenue Pierre Mendes France, F69676 Bron Cedex, France
[e] University of Clermont Auvergne, INRAE, TSCF, 63178 Aubière, France

ARTICLE INFO

Keywords:
Cultural ecosystem services
Hiking
Natural language processing
Text analysis
Wikiloc

ABSTRACT

Outdoor recreation provides vital interactions between humans and ecological systems with a range of mental and physical benefits for people. Despite the increased number of studies using crowdsourced online data to assess how people interact with the landscape during recreational activities, the focus remains largely on mapping the spatial distribution of visitors or analyzing the content of shared images and little work has been done to quantify the perceptions and emotions people assign to the landscape. In this study, we used crowd-sourced textual data from an outdoor activity-sharing platform (Wikiloc), and applied Natural Language Processing (NLP) methods and correlation analysis to capture hikers' perceptions associated with landscape features and physical outdoor activities. Our results indicate eight clusters based on the semantic similarity between words ranging from four clusters describing landscape features ("ecosystems, animals & plants", "geodiversity", "climate & weather", and "built cultural heritage"), to one cluster describing the range of physical outdoor activities and three clusters indicating hikers' perceptions and emotions ("aesthetics", "joy & restoration" and "physical effort sensation"). The association analysis revealed that the cluster "ecosystems, animals & plants" is likely to stimulate all three identified perceptions, suggesting that these natural features are important for hikers during their outdoor experience. Moreover, hikers strongly associate the cluster "outdoor physical activities" with both "joy & restoration" and "physical effort sensation" perceptions, highlighting the health and well-being benefits of physical activities in natural landscapes. Our study shows the potential of Wikiloc as a valuable data source to assess human-nature interactions and how textual data can provide significant advances in understanding peoples' preferences and perceptions while recreating. These findings can help inform outdoor recreation planners in the study region by focusing on the elements of the landscape that peoples perceive to be important (i.e. "ecosystems, animals & plants").

1. Introduction

Research on the benefits people obtain from nature has increased over the last decades (Ghermandi et al., 2023). The foremost reason is the increasing evidence that interactions with nature are indispensable for human health and well-being (Bratman et al., 2012; Soga and Gaston, 2020). Rural areas, particularly mountain landscapes, represent a vital destination for people to get close to nature, providing space for various physical and social interactions (Richins and Hull, 2016; Schirpke et al., 2018). These interactions have been conceptualized as cultural ecosystem services (CES), one of the three main ecosystem services classifications, and is defined as the non-material benefits delivered by ecosystems to humans (Haines-Young and Potschin, 2018).

Among all CESs, recreation is most noteworthy (Hermes et al., 2018), and occurs when a person is physically present in nature and interacts directly with it while engaging in physical activity (Havinga et al., 2020), whether to improve physical health through exercise or the relaxation of hiking to enjoy the landscape and its tranquility. Havinga

* Corresponding author at: University of Clermont Auvergne, INRAE, VetAgro Sup, UMR Herbivores, 63122 Saint-Genès-Champanelle, France.
E-mail address: Abdesslam.chai-allah@inrae.fr (A. Chai-allah).

https://doi.org/10.1016/j.ecoinf.2023.102332
Received 3 May 2023; Received in revised form 6 October 2023; Accepted 7 October 2023
Available online 10 October 2023
1574-9541/

et al. (2020) define this interaction as a flow of information to the person as the sensory organs decipher the immediate ecosystem configuration while carrying out physical activity. Given the importance of positive and regular interactions of people with nature to boost physical and mental well-being (Soga and Gaston, 2020), understanding the role of landscape characteristics in shaping positive perceptions and experiences for people is crucial for informed landscape management.

Traditionally, to understand people's preferences for, and perceptions of, landscape features, outdoor recreation planners and researchers often use interviews or questionnaire surveys as common methods (e.g., Howley, 2011; Plieninger et al., 2013; Wartmann and Purves, 2018). However, these methods are resource and time-intensive, often contextualized and irreproducible (Cheng et al., 2019). As a result, landscape preferences and perceptions during recreational activities remain under-explored at large spatial scales due to the lack of available information (Havinga et al., 2020).

However, recent technological advances have helped to provide access to a large amount of user-created data from social media and mobile phone applications (i.e., passively crowdsourced) that is generally inexpensive and scalable to large areas (e.g., Ghermandi et al., 2023; Ghermandi and Sinclair, 2019; Havinga et al., 2020). The wealth of information provided by these platforms through images, texts, and videos can be used to assess multiple facets of human-nature interactions (Ghermandi et al., 2023) including enjoying landscape aesthetics and tranquility (Van Berkel et al., 2018; Wartmann et al., 2021), observing wildlife (Edwards et al., 2022; Hartmann et al., 2022), and visiting parks and protected areas (e.g., Heikinheimo et al., 2017; Norman and Pickering, 2019).

Among the range of existing crowdsourcing platforms, outdoor activity-sharing platforms (e.g., GPSies, Wikiloc, and Strava) show the most promising for assessing outdoor recreational activities (Chai-allah et al., 2023; Norman and Pickering, 2019). Previous studies of recreation using outdoor activity-sharing platforms primarily focused on mapping the spatial distribution of GPS trails or analyzing the content of geolocated photographs (Callau et al., 2019; Chai-allah et al., 2023; Norman and Pickering, 2019). For example, Norman and Pickering (2019) investigated the geographical distribution of use and visitation rates in parks in relationship to park characteristics (e.g., park size, distance to urban areas and topography variables). Chai-allah et al. (2023) used GPS trails to quantify the landscape preferences of recreationists by comparing the use and potential supply patterns, while Callau et al. (2019) assessed landscape users' preferences using an automated classification of Wikiloc photographs. However, such approaches overlook the fact that landscape preferences are linked to people's emotions and perceptions (Kaltenborn and Bjerke, 2002; Wan et al., 2021).

Perceptions here can be understood as opinions, emotions or symbolic meanings that people explicitly assign to landscape characteristics to be important (Romolini et al., 2019; Huai and Van de Voorde, 2022). Such information cannot be captured using the spatial distribution of GPS trails or the content of photographs. However, it can be inferred by incorporating the textual data that can convey how people value and perceive the landscape characteristics (Fox et al., 2021a). As shown in cognitive research, statistical analyses of natural language data provide information on which entities people consider to be relevant, which is reflected in metrics such as the respective word frequencies (Günther and Rinaldi, 2022) and highlight people's representations and valuations of these entities (Günther et al., 2019).

Research has begun to explore the richness of the crowdsourced text data available on social media to study landscape preferences and perceptions, especially on urban parks (Huai and Van de Voorde, 2022; Wan et al., 2021). Several landscape features, such as water bodies, lawns, flowers and artificial elements are found to be associated with multiple park users' positive perceptions including happiness, restoration and aesthetics (Huai and Van de Voorde, 2022; Wan et al., 2021). However, how people value and perceive landscape features in rural landscapes has not been comprehensively examined. Moreover, most assessments focus on generic groups of recreationists (e.g., naturalists, photographers, and sports) (e.g., Chai-allah et al., 2023), but little has been done by specifying a user group (e.g., hikers, bikers, and runners). Furthermore, while automatic image classification through pre-trained image recognition tools (e.g., Google Cloud Vision and Clarifai) is largely increasing in the literature (e.g., Lee et al., 2019; Fox et al., 2021a; Egarter Vigl et al., 2021), text-based studies have commonly been relying on manual methods (Calcagni et al., 2022; Schirpke et al., 2021; Wan et al., 2021). Overall, these studies have shown that manual coding can be effective especially if the person coding is familiar with the area, but on the other hand, it keeps the processes laborious specifically in data filtering given the unstructured nature and the immense volume of crowdsourced text data available on social media.

Computational methods such as natural language processing (NLP) techniques are alternative approaches to 'mine' recreationists' emotions toward nature from crowdsourced text (Schirpke et al., 2023). These methods can offer effective tools for the transformation of enormous amounts of unstructured crowdsourced text into structured and systematic information (Hirschberg and Manning, 2015; Huai and Van de Voorde, 2022; Purves et al., 2022). Specifically, word embedding methods including Latent Semantic Analysis (LSA), word2vec and GloVe (Mikolov et al., 2013; Naili et al., 2017), can build a low-dimensional vector representation of words from a collection of text to grasp the semantics of the respective words. Words with similar meanings will tend to be located near each other in the vector space, allowing the computation of similarity between the words and therefore creating a set of topics that englobe words with similar meanings (Edwards et al., 2022; Huai and Van de Voorde, 2022; Mikolov et al., 2013). A scarce body of research has started to combine crowdsourced text from social media with NLP methods to capture the landscape characteristics mentioned by recreationists during their outdoor activity and how they value and perceive these landscape features (Fox et al., 2021a; Gugulica and Burghardt, 2023; Huai and Van de Voorde, 2022). For example, Huai and Van de Voorde (2022) used Word2vec and online reviews to gain insights into how people perceive environmental features in Shanghai and Brussels urban parks, demonstrating the suitability of NLP techniques to understand people's perceptions of the landscape. Nevertheless, studies show that relying only on automatic classifications without manual checking may lead to an inaccurate understating of how nature benefits people considering the polysemy of words and the diverse modes of expressing perceptions (Ghermandi and Sinclair, 2019).

Therefore, in this work, we adopted a fully featured crowdsourced data-driven methodology and semi-automatic approach to answer the following questions: 1) What are the landscape features and outdoor physical activities preferred and mentioned by hikers? 2) What are the perceptions mentioned by hikers? 3) What perceptions do hikers associate with landscape features and outdoor physical activities?

2. Materials and methods

Our study aims to understand hikers' preferences and perceptions during an outdoor experience using crowdsourced textual data from Wikiloc. To achieve our goal, we used the Auvergne region in France as a case study. Auvergne is a mountainous region with volcanic massifs ranging from 159 to 1874 m above sea level. The 26,000 km^2 region is covered by a mix of cropland, forests and grassland livestock farming systems. Auvergne is mostly rural, with urban areas only representing 3% of its area (Corine Land Cover, 2018: https://land.copernicus.eu/pan-european/corine-land-cover/clc2018). It also contains four regional nature parks, a UNESCO heritage site and diverse archaeological and historical sites (Supplementary Materials, Fig. S1). These various biophysical and cultural characteristics make Auvergne an attractive region for hiking.

We adopted a methodology based on a data-driven approach (Sebei et al., 2018) consisting of three main steps: data collection, data processing and data analysis (Fig. 1). The collection of Wikiloc data in Auvergne was done in a Python environment (script available at this link: https://github.com/achaiallah-hub/Wiki4CES). Data processing, analysis and charts were done in R, version 4.1.0 (R Core Team, 2021).

2.1. Data collection

Wikiloc offers over 14 million hiking trails and has increasing popularity with the number of users increasing from 3.5 million in 2019 to over 11.3 million in 2023 (https://www.wikiloc.com). The Wikiloc platform allows users to record and share outdoor trails, providing associated textual descriptions, photographs, and their spatial location (latitude and longitude) on the trail. The type of outdoor activity is also recorded (hiking, cycling, or running), as well as the date the trail was made. Wikiloc has a broad user base and potential as a data source for diverse CES (Callau et al., 2019; Chai-allah et al., 2023; Norman and Pickering, 2019). While the trails and photographs from this platform have been used in CES studies (Callau et al., 2019; Chai-allah et al., 2023; Norman and Pickering, 2019), the textual data remained unexplored.

Wikiloc trails available in Auvergne were selected for the years 2017–2020 resulting in 1128 trails uploaded by 446 users. We pre-filtered the data by 1) keeping only the trails that were assigned by users as hiking trails, 2) removing those without associated text data, and 3) applying the "Trail-User-Days" (TUD) measure (i.e., a single trail per user per day), introduced by Chai-allah et al. (2023) to filter the social media data to represent a day trip. After pre-filtering 978 hiking trails from 388 users remained for further use (see three examples presented in Table S1 in the Supplementary Materials). Finally, the data sets were anonymized and any unnecessary metadata was removed.

2.2. Data processing

To identify the landscape features, physical outdoor activities, and the perceptions mentioned by hikers during their outdoor experience, we extracted the most frequent single words (unigrams) from Wikiloc posts. The extraction of unigrams consisted of three steps: text translation, data pre-processing, and word frequency counting. First, since the text description included five different languages (French, English, Dutch, Italian, and Spanish), we translated all text descriptions into English using DeepL (https://www.deepl.com). Second, data pre-processing consists of using natural language processing (NLP) methods to perform automated text mining in R, primarily using the packages tidytext (Silge and Robinson, 2017) and tm (Feinerer and Hornik, 2018) for tokenization, lemmatization and English stop words removal based on a predefined list of common English words (e.g., and, the). An additional set of words such as the name of the country, towns, places, and those related to transportation (e.g., car, parking, and train) were also removed for the lack of relevance to the purpose of the study. Third, we measured the frequency of the extracted words, and only those occurring at least three times in total were selected for further analysis (Huai and Van de Voorde, 2022; Schirpke et al., 2021). Finally, to understand which words were more typical in our corpus compared to a general corpus, we compared the frequency distribution we observed in the Wikiloc corpus with the frequency distribution in the general web-crawled ukWac corpus (Baroni et al., 2009) using a chi-square test by focusing on the individual contribution of words to the total effect size (Oakes and Farrow, 2006).

In addition, to ensure that the extracted single words were used in a positive context indicating a CES (Fox et al., 2021a) and to count them only when they appear in a positive context such as "good hike" and not in case of a negative context like complaining about a "boring hike", the following steps were followed. First, we extracted the bi-grams i.e. pairs of consecutive words in a given sample of text that helped to identify how words are used in combination (e.g., Hausmann et al., 2020). Second, we considered only the bi-grams that include at least one of the extracted high-frequency words (Supplementary Materials, Fig. S4). Third, we manually classified bi-grams into positive or negative

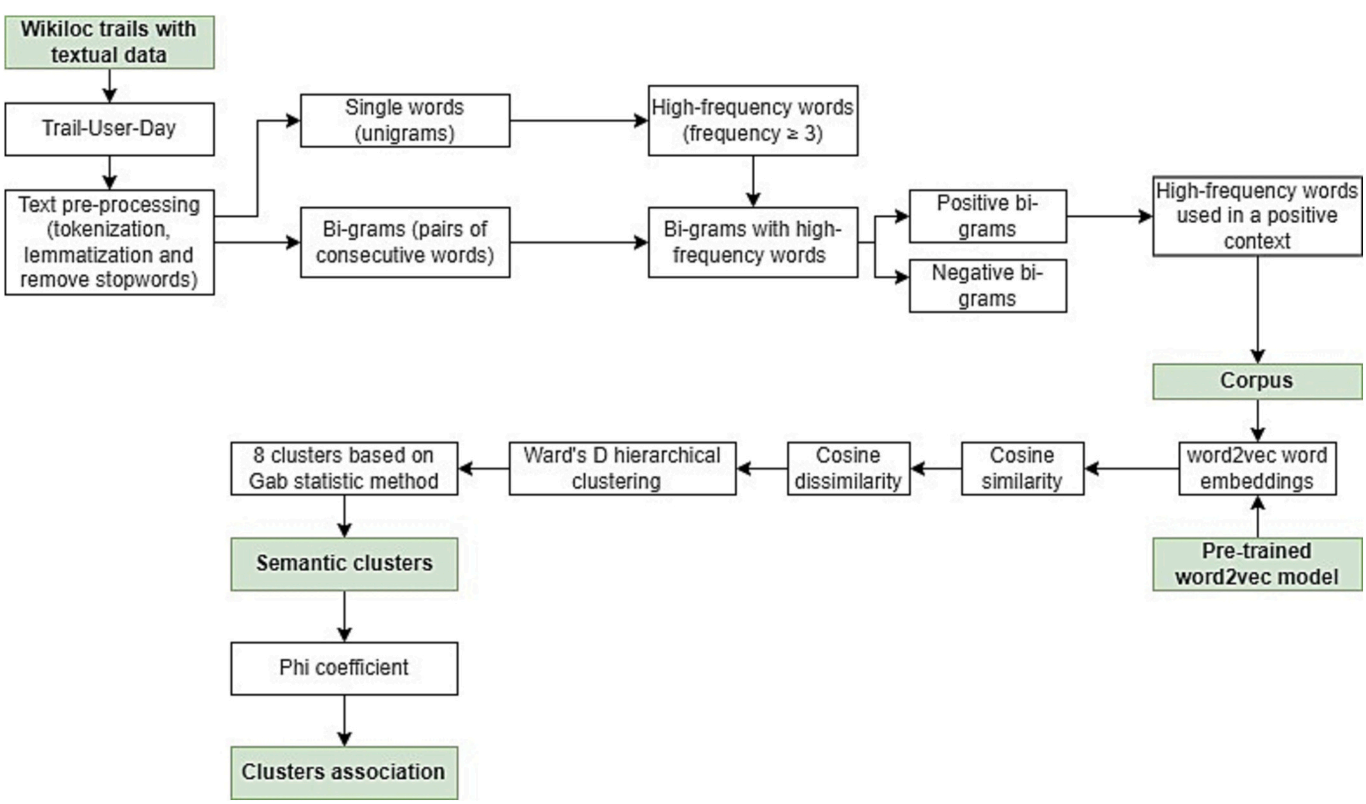

Fig. 1. Workflow chart depicting methodological steps of this study.

associations following Lampinen et al. (2021), for example, bi-grams such as “good hike” were classified as positive associations and bi-grams such as “boring hike” as negative associations. Finally, we recalculated the frequency of words by considering only when they appear in a positive context. However, as all selected words only appeared in a positive context their frequencies did not change.

2.3. Data analysis

2.3.1. Semantic clustering

To identify the main topics mentioned by hikers during their outdoor experience, we classified the selected unigrams into structured clusters based on semantic similarities between words. To do this, we used word embeddings, an NLP method where machine learning algorithms represent words as distributional vectors (e.g., Levy et al., 2015; Mikolov et al., 2013), based on the concept that semantically similar words are used in similar linguistic contexts (i.e., they have a similar distribution over these contexts; Sahlgren, 2008). The similarity between two vectors can be computed using algebraic metrics such as the cosine between these vectors. For example, words like “waterfall” and “lake” have similar word vectors to the word “river” but distinct ones from the word “church”.

Word2vec is a widely known word embedding technique (e.g., Mikolov et al., 2013; Naili et al., 2017), and has been a successful tool for investigating CES from social media data (e.g., Gugulica and Burghardt, 2023). Word2vec can be carried out in two ways: by creating and training your own word2vec space based on two different architectures, namely, Skip-gram, which predicts the target word based on a nearby word, and Continuous Bag-of-Words (CBOW), which learns to predict a target word according to its context (Mikolov et al., 2013), or by using pre-trained models (e.g., Baroni et al., 2014). Training a new custom word2vec model requires a very large dataset, increases the computational effort, and could be biased by the geographic origin of the data. Therefore, we performed the semantic analysis using the pre-trained *baroni* word2vec model provided at http://www.lingexp.uni-tuebingen.de/z2/LSAspaces (Günther et al., 2015). This model has been demonstrated to generate high-quality performance with dense word vectors and to produce the best results regarding semantic similarity tasks (Baroni et al., 2014). This space was created using the CBOW algorithm and contains vectors for 300,000 different words, covering a broad variety of different topics. It was trained from a 2.8-billion-word corpus, a concatenation of the ukWaC corpus (web pages material from . uk domain; Baroni et al., 2009), Wikipedia, and the British National Corpus (BNC Consortium, 2007).

The cosine similarities of the extracted high-frequency words from our corpus were then carried out using the *baroni* word2vec space and the LSAfun package in R (Günther et al., 2015). The package was originally written for LSA, but can be used with all vector space models and its combination with pre-trained spaces was reported to produce adequate results (Günther et al., 2015). Cosine similarity is measured as the cosine of the angle between vectors, with values that range from 0 to 1, with 1 indicating total similarity and 0 representing total dissimilarity. Finally, the similarity matrix was transformed into a distance matrix by calculating the cosine dissimilarity (1 - cosine similarity) in order to be used in hierarchical clustering. The hierarchical clustering was made using Ward's distance method and the fastcluster R package (Müllner, 2013). The optimal number of clusters was determined by the Gab statistic method (Tibshirani et al., 2001), indicating $k = 8$ as an optimal number of clusters (Supplementary Materials, Fig. S2).

The results of word clustering were manually verified to ensure the accuracy of the clusters and the list of words per cluster was used as a basis for labeling the clusters following the existing literature on CES and human-nature interaction (Kaiser et al., 2021; Egarter Vigl et al., 2021; Wan et al., 2021; Huai and Van de Voorde, 2022; Gugulica and Burghardt, 2023). Finally, we counted the frequency of each cluster using the count function in R. Following Wan et al. (2021), we considered each cluster's frequencies as hikers' revealed preferences.

2.3.2. Clusters association

To examine the perceptions that hikers associate with landscape features and physical outdoor activities, the Phi coefficient was carried out through the widyr package in R (Silge and Robinson, 2017). The Phi coefficient is a measure of binary correlation between two categorical variables, which indicates how often they appear together relative to how often they appear separately. Here, it was used to indicate the number of documents (the number of trails) where, for example, clusters A and B appear, or neither do, compared to the number of documents where one appears without the other, considering eq. 1:

$$\phi = \frac{n_{11}{}^{*}n_{00} - n_{10}{}^{*}n_{01}}{\sqrt{(n_{1.}{}^{*}n_{0.}{}^{*}n_{.0}{}^{*}n_{.1})}} \tag{1}$$

Where n_{11} is the number of documents with both clusters A and B, n_{00} is the number of documents with neither clusters A and B, n_{10} is the number of documents with cluster A but without cluster B, n_{01} is the number of documents with cluster B but without cluster A, $n_{1.}$ is the number of documents with cluster A, $n_{0.}$ is the number of documents without cluster A, $n_{.1}$ is the number of documents with cluster B, and $n_{.0}$ is the number of documents without cluster B.

All correlations are presented following the Quinnipiac University scale (Akoglu, 2018), therefore, correlations below 0.05 were considered weak, and those above 0.25 as very strong.

3. Results

3.1. Data collection and preprocessing

A total of 978 hiking trails were associated with textual data, corresponding to 86.7% of the collected data, with an average text description length of 14 words (SD = 28.6). Half of the trails were recorded in the summer months from June to August (481 trails), and 34% of the trails fell on weekends (Supplementary Materials, Fig. S3). The maximum number of trails uploaded by a single user was 30, while 207 users uploaded only one trail over the four-year study period. Moreover, 183 unique words occurred at least three times in total and in a positive context, with a cumulative frequency of 3553. The words “view”, “trail” and “walk” were the most frequent unigrams and no negative words such as “sad” or “boring” were found in the final dataset (Supplementary Materials, Fig. S4). There was a significant difference in the contents between the two corpora Wikiloc and ukWaC ($x^2 = 1{,}520{,}366$, df = 17,317, $p < 0.001$), with words such as “hike” and “trail” appearing more frequently in the Wikiloc corpus compared to ukWaC corpora, reflecting the recreation-oriented use of Wikiloc (Supplementary Materials, Fig. S5).

3.2. Clustering and labeling

We identified eight different clusters based on 183 unique words, with a cumulative frequency ranging from 91 to 805 (Fig. 3). Four out of the eight clusters were related to landscape features (41%; cumulative frequency of the four clusters divided by the total frequency of the eight clusters), one described physical recreation (23%) and three were related to user perceptions (36%). Regarding the clusters describing the landscape features, which we named “geodiversity”, “ecosystems, animals & plants”, “built cultural heritage” and “climate & weather”, hikers mentioned natural elements (78%) more than non-natural elements (22%) with the predominance of the geodiversity cluster on natural elements (63%). Words related to terrestrial outdoor activities such as “trail”, “hike”, and “walk” were more frequent than words describing aquatic activities in the cluster “outdoor physical activities”. Moreover, the most frequent words in the cluster “ecosystems, animals & plants” were related to ecosystems (e.g., forest and grasslands) rather than specific animals and plants (both of which appear with low occurrences; Supplementary Materials, Table S2). The cluster "geodiversity" showed a variety of elements ranging from geomorphological features (e.g., valley, volcano, and mountain) to hydrological features (e.g., lake, waterfall, and river), while the cluster "climate & weather" was reflecting the day-time recreational use of the area with words such as "sunshine" and "shadow". The non-natural features related to attractions (e.g., castle and church) were the most popular in the cluster "built cultural heritage" (Fig. 2).

Of user perception clusters, the "aesthetics" cluster is the most frequent (51%), followed by "joy & restoration" (33%) and "physical effort sensation" (16%) (Fig. 3). The cluster "aesthetics" is defined by the frequent use of words related to visual beauty such as "view", "beautiful" and "panoramic". The cluster "joy & restoration" showed a variety of words, ranging from words about the quality of the experience such as "nice" and "good" to words about the restorative experience (e.g., break, rest, and relax). Finally, the cluster "physical effort sensation" comprises words such as "easy" and "hard" (Fig. 2).

3.3. Association between landscape features, physical recreation and hikers' perceptions

The association analysis revealed hikers' perceptions of landscape features and physical recreation (Fig. 4). The clusters "ecosystems, animals & plants" and "climate & weather" were associated with the three hikers' perceptions ("aesthetics", "joy & restoration" and "physical effort sensation"), with the "aesthetics" perception being more associated with the cluster "ecosystems, animals & plants". The "geodiversity" cluster was only associated significantly with "physical effort sensation" perception, while the results show weak associations between "built cultural heritage" and hikers' perceptions. Finally, hikers strongly associate the cluster "outdoor physical activities" with two perceptions ("joy & restoration" and "physical effort sensation").

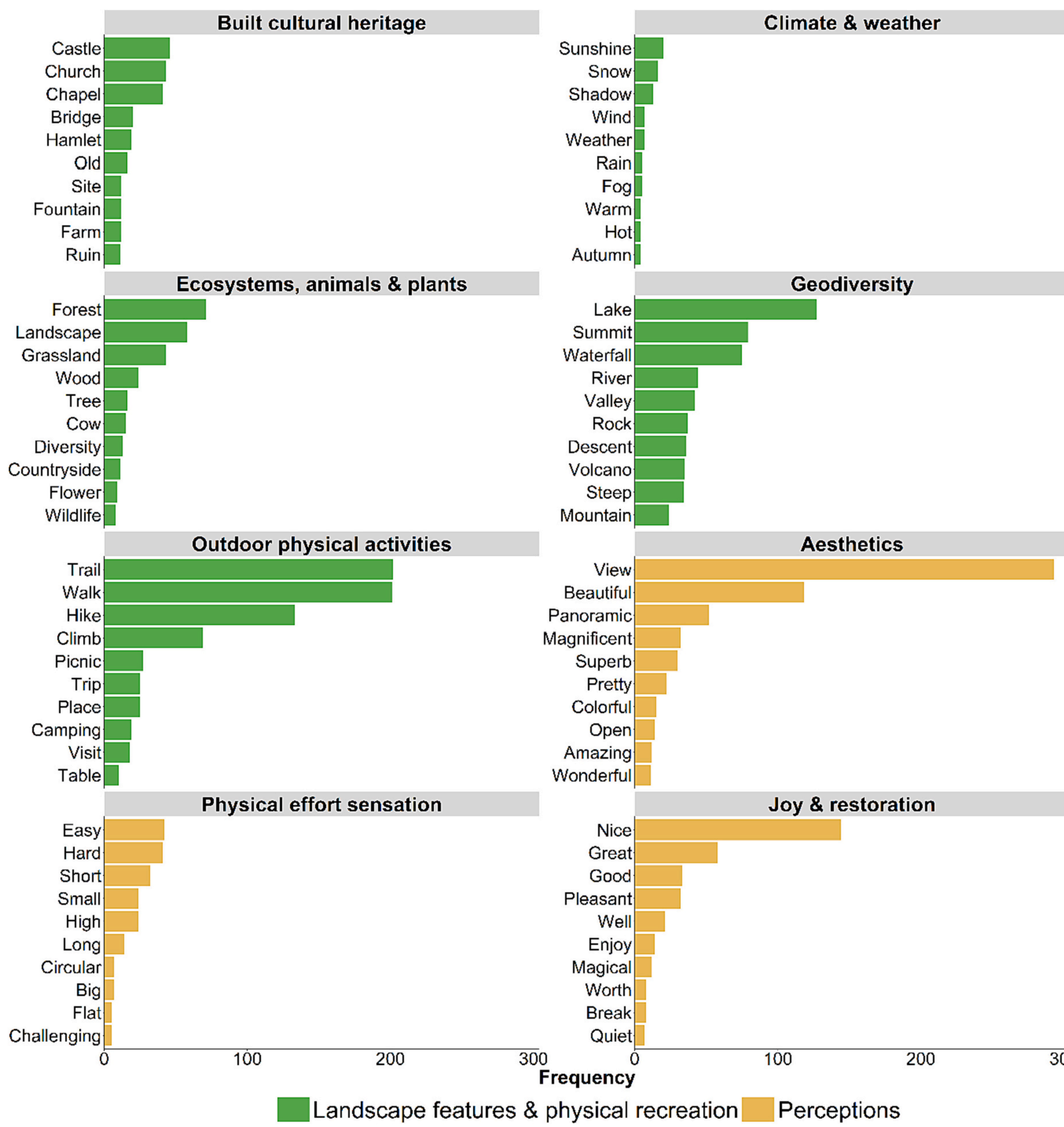

Fig. 2. Top ten words of each identified cluster and their frequencies of occurrence. Complete details on the clustering results can be found in Table S2 in the Supplementary Materials.

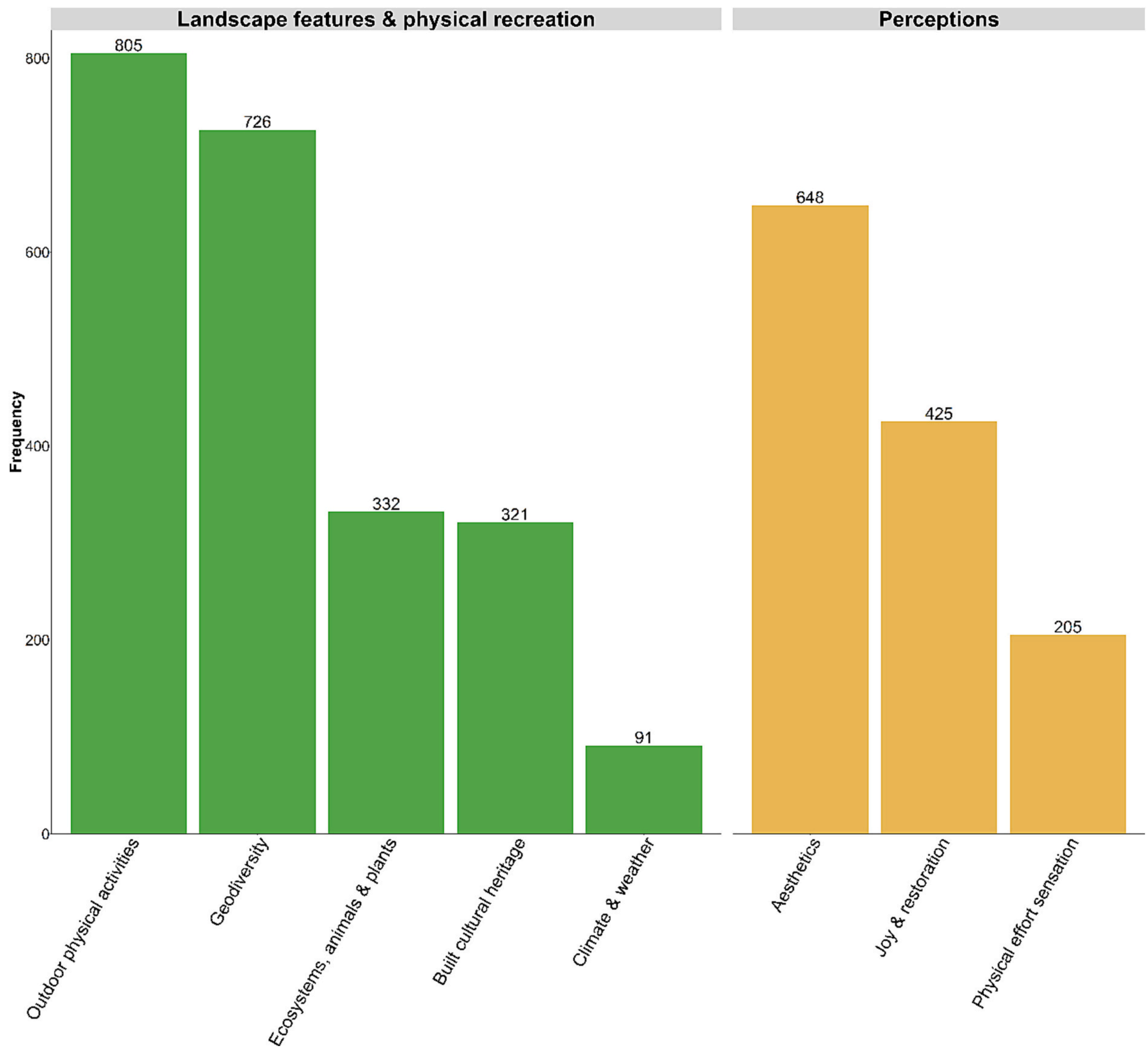

Fig. 3. Frequency of occurrence of landscape features and physical recreation clusters (left) and perceptions clusters (right).

4. Discussion

4.1. Preferences and perceptions of hikers in a rural landscape

Overall, for the preferences, we found four clusters describing the landscape's biophysical features and one cluster describing the range of physical outdoor activities. Among the biophysical features, people mentioned geodiversity features more than ecosystems, animals and plants features. This can be explained by the fact that Auvergne is characterized by its iconic geodiversity landscapes related to its geological history, which leads to 80 volcanoes and multiple mountains and hills, ensuring the development of geotourism in the area (Cayla, 2014). Moreover, geodiversity features such as volcanoes could gain greater attention from visitors to be visited and shared on social media as impressive and unusual structures that symbolize the experience in Auvergne. Other studies have identified similar patterns where mountains, lakes and rivers are popular landscape features to be visited during recreational activities and then posted on social media (e.g., Pickering et al., 2020; Schirpke et al., 2021). The height of mountains is particularly associated with hiking (Aiba et al., 2019). It can thus explain why the word "summit" received a high number of mentions in this study as hikers may relate to demonstrating they were at the highest point of the mountain when it is harder to get there.

The cluster "ecosystems, animals & plants" is largely composed of words describing ecosystems (e.g., forest and grassland) and livestock breeds such as "cows" rather than those describing specific wild animals, plants and flowers. This may refer to hikers wanting to show their familiarity with local breeds as part of the rural identity of the region (Barry, 2014), together with the breeds' animals being easily noticeable by the large public compared with wild animals, plants, and flowers. However, the use of words such as "flower" and "wildlife", even though they are less frequent, emphasizes a desire for wild nature.

Hikers mentioned elements of the "built cultural heritage" cluster almost as much as "ecosystems, animals & plants" cluster elements. These results echo the viewpoint that human-made landscape features mainly those that relate people to the past and history of the area (e.g., church or castle), can draw visitors' attention and enhance landscape experiences (e.g., Calcagni et al., 2022; Van Berkel et al., 2018). Moreover, the cluster "outdoor physical activities" was not solely focused on the physical interactions with nature (e.g., hiking, walking, and climbing) but also included elements that describe social relations (e.g., picnic, camping, and family), which implies that people give meaning to both aspects when interacting with nature. These results corroborate the findings of a survey study in Germany that indicated that rural landscapes are multifaceted landscapes that offer space for various types of human-nature interactions (Plieninger et al., 2013).

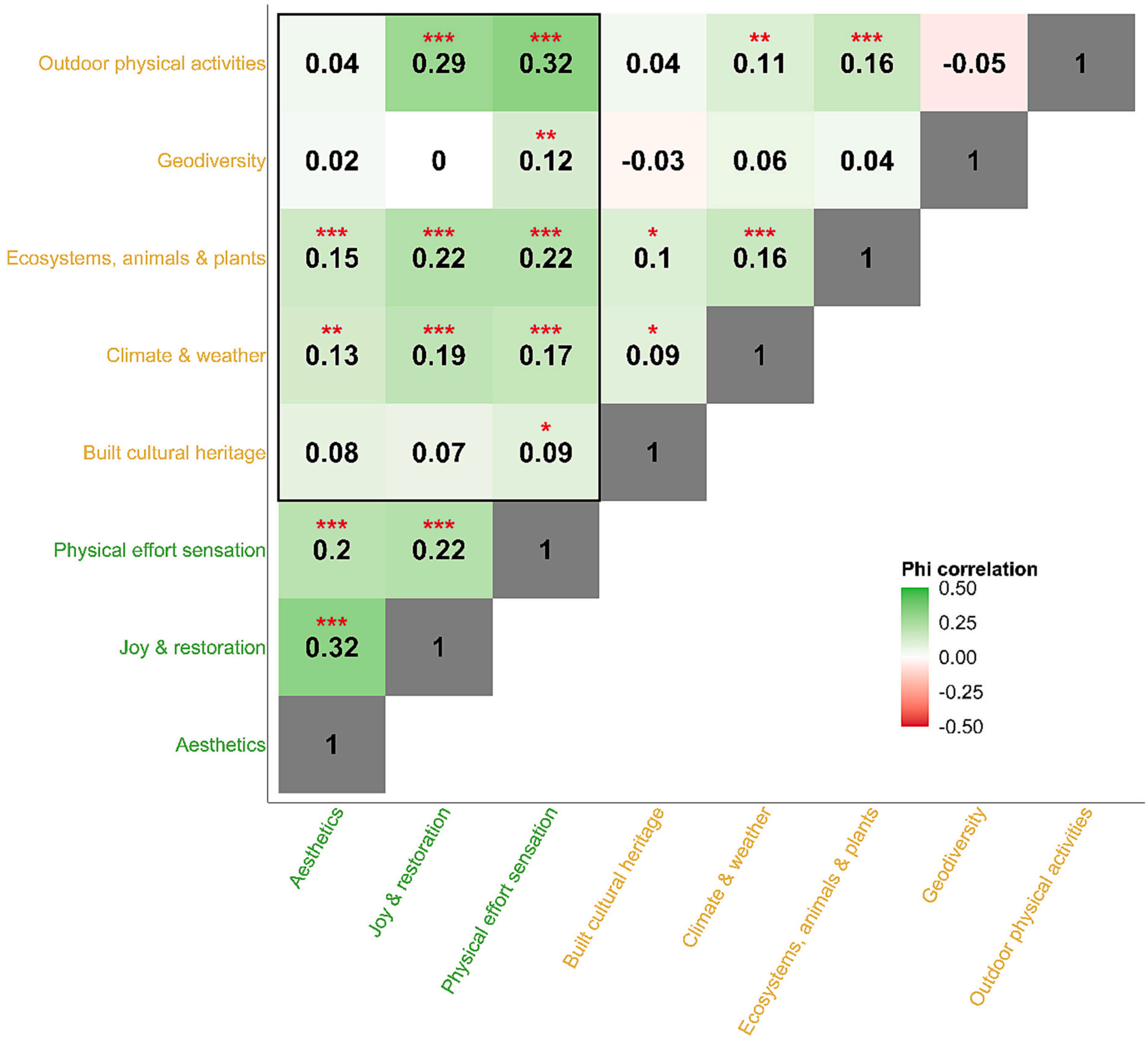

Fig. 4. Associations correlogram. The black rectangle shows the studied correlations between landscape features, physical recreation (“built cultural heritage”, “climate & weather”, “ecosystems, animals & plants”, “geodiversity” and “outdoor physical activities”) and hikers' perceptions (“aesthetics”, “joy & restoration” and “physical effort sensation”). For significant correlations, * indicates *p*-value <0.05, ** indicates p-value <0.01, *** indicates p-value <0.001.

Overall, the aesthetic expressions lead to the most frequent perceptions that hikers allocated to the landscape, and were mostly expressing visual beauty such as “view”, “beautiful” and “panoramic”. This result is in line with interview-based studies where aesthetic perception is a frequent response to landscape environments (Gobster and Westphal, 2004; Kaiser et al., 2021), especially in mountainous areas where long and open views supply higher aesthetic values mainly on sunny days (Pickering et al., 2020; Schirpke et al., 2018). Our findings indicate that hikers express broad emotions, mainly joy, and restoration, as well as the physical effort perceived during their outdoor experiences in nature. Where the landscape was perceived as “nice” and “great” as well as having space for “rest” and “break”, seeing and being in nature can promote positive emotions and stress reduction for visitors as an escape from stressful urban lifestyles (Hansmann et al., 2007; Hussain et al., 2019). Alternatively, visitors have considered the physical effort engaged during recreation, which may indicate visitor motivations for hiking. For instance, some hikers may choose close and challenging tracks as part of the daily use of the surrounding nature for exercise purposes, whereas hikers with children may prefer small tracks for easy walks with families that are less physically challenging (Wilcer et al., 2019).

These insights reveal 1) what hikers prefer and consider worth sharing on social media and how they describe their perceptions during an outdoor experience and 2) the richness of Wikiloc textual data, which allows for spotting a richer range of hikers' preferences for natural and non-natural physical elements of the landscape, as well as their perceptions that have been scarcely studied in the current literature.

4.2. *Hikers' perceptions associated with outdoor physical activities and landscape features*

The association analysis revealed many positive and significant correlations between hikers' perceptions of outdoor physical activities and landscape features. This result provides evidence of the “bundled” nature of how people perceive the landscape while engaging in recreational activities. People generally do not assign specific perceptions to specific landscape components, which aligns with a previous study based on interviews (Plieninger et al., 2013). This suggests that people do not perspicuously distinguish one cultural service gained from nature from the other, but enjoy them in a bundle. Here, our results add more details by focusing on hikers as a target user group, which is an important contribution of this study since it provides more specific information by the type of recreational activity, as suggested by several studies that use social media to understand human-nature interactions (e.g., Lee et al., 2019). However, to the best of our knowledge, this has never been done before.

In addition, by exploring these associations, we can begin to better understand the relative importance of some landscape features. Here, the cluster “ecosystems, animals & plants” was mentioned less frequently than the cluster “geodiversity”. The hikers do however associate it with more perceptions than the cluster “geodiversity”, mainly “joy & restoration” and “aesthetics”. This may be related to the fact that Auvergne is a mountainous region and mountains are common while encountering unexpected wildflowers in grasslands, flowering trees, or refreshing forests may lead to more visitors' appreciation. These findings are consistent with a recent study conducted on urban parks in Hong Kong, which showed that park visitors associate flowers with aesthetic value, and trees and lawns with restorative value (Wan et al., 2021). Moreover, being surrounded by a calm landscape like grasslands has been found to evoke tranquility and relaxation in people (Wartmann and Purves, 2018). Although lakes and waterfalls were highly mentioned, the prohibition of water-based activities in many of them in Auvergne may have led visitors to less associate them with particular perceptions. The search for a challenging landscape with high geomorphological variation in terms of slope and elevation may explain the association of the cluster “geodiversity” with physical effort sensation as Wikiloc users are more likely to prefer sites with rugged terrain (Chai-allah et al., 2023; Norman and Pickering, 2019).

Physical activities in nature contribute considerably to health and well-being benefits for individuals (Hansmann et al., 2007; Wolf and Wohlfart, 2014), this may explain why hikers strongly associate the cluster “outdoor physical activities” with the perceptions of “joy & restoration” and “physical effort sensation”. Moreover, hiking expectations and motivations vary among individuals; for instance, experienced hikers may be more satisfied by difficult hikes and less aware of the intensity of their physical activity (Chhetri, 2015; Wolf and Wohlfart, 2014), whereas less experienced hikers or those with young children can find easy hikes more suitable and enjoyable. This suggests that how hikers perceive the site or the landscape may depend on their motives for the activity of interest, and thus human-nature relationships are complex and context-dependent (Fox et al., 2021a).

The analysis of associations is one of the advancements of our study compared to previous studies using text data social media to capture human-nature interactions (e.g., Pickering et al., 2020; Schirpke et al., 2021). These studies usually consider what is mentioned as the most important landscape features for people. However, our results highlight an important discrepancy, as what is mentioned the most. Here “Geodiversity” is mentioned the most in posts, but does not have the highest correlation with visitor perceptions, with “Ecosystems, animals and plants” having the highest. Thus, simply relying on frequency counts - how often something is mentioned - is not by itself sufficient to fully understand to what extent people appreciate and value certain features. By enriching the analysis with the perceived perceptions of visitors presented here or by sentiment analysis (Fox et al., 2021a), we can begin to better capture the importance of specific physical features in ways that go beyond purely measuring their number of occurrences. This approach helped us to 1) disentangle the components aimed during a recreational experience through the perception clusters and 2) assess the element of landscapes that provide them through the natural features' clusters.

4.3. Wikiloc as text-based social media

The current study suggests Wikiloc as a rich source of textual data to study human-nature interactions. The platform has the potential to solve the issue of data availability rising with the decreasing popularity of Flickr, the principal data source for social media-based studies, restrictive access to data for research purposes (like in Instagram and Twitter), or even ceased operating (e.g., Panoramio) (Ghermandi et al., 2023). Moreover, Wikiloc textual data presents multiple advantages over most text-based social media sites by being geolocated and without a character limit (e.g. Reddit posts are not geolocated and Twitter has a 280-character limit) (Fox et al., 2021b). This has allowed us to assess a large spectrum of hikers' preferences and perceptions.

The presented approach of mining crowdsourced text has in addition major advantages compared to examining the visual content of shared images as it allows access to feelings and perceptions that hikers associate with landscape features and outdoor experiences, which are impossible to obtain from images. Calcagni et al. (2022) found that textual data allow for spotting a wider variety of CES than could be achieved solely through visual data analysis. Moreover, by using NLP methods to filter the words and to analyze how the identified single words are used in combination (bi-grams), we ensured that they provide a positive benefit indicating a CES, which was one of the drawbacks of the studies using single-word classification. Besides, the approach of clustering is based on pre-trained word2vec models, which allowed us to rapidly classify words with similar meanings into topics that reflect a wider range of hikers' preferences and perceptions. By exploring Wikiloc we reveal the unique opportunity it offers by bringing uniquely rich data to assess multiple aspects of human-nature interactions, which can help overcome the existing gap in data available for researchers working on CES assessment using social media.

4.4. Implications for rural landscape management

With the increased use of rural landscapes for multiple recreation activities, exploring the perceptions people assign to landscape features can help outdoor recreation planners adjust policies based on what people find to be important. We reported that the cluster describing ecosystems, animals and plants had the highest association with visual aesthetic perception. This suggests that hikers seek out open and natural landscapes while hiking. Land managers in Auvergne should focus on maintaining this openness as it seems to be a key factor for hikers' well-being. Additionally, we found less association between geodiversity features such as lakes and rivers with hikers' perceptions, underlying a need for future improvements and restoration regarding these blue areas, for example increasing their accessibility or water quality. Previous research has shown that restoration management has a sound potential to improve people's benefits from blue areas (Kaiser et al., 2021). Our findings on the bundled nature of hikers' perceptions of the landscapes can also serve as a baseline for creating new trails that provide access to diverse and appealing landscapes but are also challenging in terms of physical effort as an opportunity to promote physical well-being.

4.5. Limitations and future research

Although this study demonstrates that hikers' preferences and perceptions can be achieved by combining textual data from Wikiloc and NLP methods, several caveats need to be considered. First, demographic and sample biases with the socio-demographic information of social media users are never entirely clear, difficult to obtain and can change following the popularity of the platforms, which keeps the question of data representativeness challenging. In a recent study, Venter et al. (2023) found that despite the increased proportion of total recreationists that Strava–an outdoor activity-sharing platform similar to Wikiloc–represents over time, it is biased toward higher-income populations, males, and middle-aged people between 35 and 54 years. Such an analysis must be extended to Wikiloc and considered in future studies. Our results should therefore not be generalized as such to a wider population.

Second, we have limited our analysis of anthropogenic features to cultural-historical elements, considering that are key factors for visitation hotspots and for their cultural and religious values (Calcagni et al., 2022; Van Berkel et al., 2018). However future research could expand our understanding of recreationists' perceptions to other anthropogenic properties such as features of accessibility and touristic facilities that were not investigated in this study, considering that people can only enjoy specific landscapes when they have access to it (Costanza, 2008).

Third, we performed word clustering using a word2vec space trained on a large web corpus. Although this model gives satisfactory results and is expected to effectively scale to larger areas, we were unable to use it without translating non-English texts into English as it was trained on an English corpus. While this automatic translation might have enabled us to unify the dataset and have more compact clusters considering the multiple languages found in the original dataset, it may have masked subtle linguistic delicacies and metaphorical expressions in different languages. Future studies would do well to include the linguistic distinctions between social media users so that we can explore how different language communities perceive landscape features during recreational activities. Yet we need to emphasize that word-sense ambiguity issues caused certain misclassification but have occurred very sporadically and were easily noticeable in the manual checking. However, these NLP algorithms are expected to improve in the future, primarily by including textual data from social media (Gugulica and Burghardt, 2023) or applying the emerging transformer-based models, for example, the Bidirectional Encoder Representations from Transformers (BERT) that have been shown as a promising tool for text classification (Devlin et al., 2018; Hunter et al., 2023).

Fourth, while social media data has the advantage of being spontaneously produced, people are more likely to share what they like and think is worth sharing, resulting in more sharing of positive experiences over negative ones (Hausmann et al., 2020; Wan et al., 2021), and hence creating an overall positivity bias and incomplete image of human-nature interactions. Future research should combine crowdsourced data analysis and online surveys with users of the studied platforms to get a more exhaustive and complete understanding of human-nature interactions.

Finally, we used only one outdoor activity-sharing platform, Wikiloc, and focused on one recreational activity, hiking. This potential bias can be minimized by including other platforms and exploring how other user groups such as runners or bikers interact and perceive landscape features. In addition, it is also known from other studies that CES are context-dependent and therefore our findings need to be understood considering the context studied here (i.e. rural landscapes).

5. Conclusion

Our work complements previous research with three major contributions. First, we analyzed hiking data from Wikiloc, a crowdsourced sports platform, which provides access to a more homogenous segment of the population because users share similar interests in outdoor activities. It offers the opportunity to assess how one or multiple user groups of recreationists interact with nature which is hardly possible in other common platforms used in CES assessment such as Flickr. Second, we explored textual data that allows access to users' feelings and perceptions, which were scarcely studied in the literature as the focus was on image content. This process has the potential to enhance the current understanding of what landscape features mean for people by identifying those that people find to be important and therefore to be considered in landscape management. Third, we provided a semi-automated, flexible and transferable data-driven approach using NLP techniques that can overcome some of the limitations of manual methods.

Our results suggest that natural features of the landscape are more likely to stimulate different perceptions of hikers from aesthetics, joy, and restoration to the physical effort sensation during hiking. These perceptions could be independently obtained in urban areas, for example in museums, sports rooms, or by meeting friends in coffee. Through our results, we found that natural features - mainly the cluster 'Ecosystems, animals & plants' - stimulate the three perceptions together. This suggests that hikers go to rural landscapes to experience "Aesthetics", "Physical effort and sensation", and "Joy & restoration", because they offer them these perceptions in a bundle, i.e. at the same time. This result can help to understand the implications of accessible hiking trails from urban centers in terms of human well-being and health.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

I have shared the code in the manuscript

Acknowledgements

This work was financially supported by the French government's national research agency (ANR) and its "Investissements d'Avenir" program (16-IDEX-0001 CAP 20-25). NFs contribution to the manuscript was supported by the Eric and Wendy Schmidt AI in Science Postdoctoral Fellowship, a Schmidt Futures program. FGs contribution was supported by DFG Emmy-Noether grant "What's in a name?" (project nr. 459717703). We thank Anne de la Foye for her help with R and Valentin H. Klaus for commenting on an earlier version of this manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ecoinf.2023.102332.

References

Aiba, M., Shibata, R., Oguro, M., Nakashizuka, T., 2019. The seasonal and scale-dependent associations between vegetation quality and hiking activities as a recreation service. Sustain. Sci. 14, 119–129. https://doi.org/10.1007/s11625-018-0609-7.

Akoglu, H., 2018. User's guide to correlation coefficients. Turk. J. Emerg. Med. 18, 91–93. https://doi.org/10.1016/j.tjem.2018.08.001.

Baroni, M., Bernardini, S., Ferraresi, A., Zanchetta, E., 2009. The WaCky wide web: a collection of very large linguistically processed web-crawled corpora. Lang. Resour. Eval. 43, 209–226. https://doi.org/10.1007/s10579-009-9081-4.

Baroni, M., Dinu, G., Kruszewski, G., 2014. Don't count, predict! A systematic comparison of context-counting vs. context-predicting semantic vectors. In: Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Association for Computational Linguistics, Baltimore, Maryland, pp. 238–247. https://doi.org/10.3115/v1/P14-1023.

Barry, S.J., 2014. Using social media to discover public values, interests, and perceptions about cattle grazing on park lands. Environ. Manag. 53, 454–464. https://doi.org/10.1007/s00267-013-0216-4.

BNC Consortium, 2007. The British National Corpus, Version 3. Available at. http://www.natcorp.ox.ac.uk/.

Bratman, G.N., Hamilton, J.P., Daily, G.C., 2012. The impacts of nature experience on human cognitive function and mental health: nature experience, cognitive function, and mental health. Ann. N. Y. Acad. Sci. 1249, 118–136. https://doi.org/10.1111/j.1749-6632.2011.06400.x.

Calcagni, F., Nogué Batallé, J., Baró, F., Langemeyer, J., 2022. A tag is worth a thousand pictures: a framework for an empirically grounded typology of relational values through social media. Ecosyst. Serv. 58, 101495. https://doi.org/10.1016/j.ecoser.2022.101495.

Callau, A.À., Albert, M.Y.P., Rota, J.J., Giné, D.S., 2019. Landscape characterization using photographs from crowdsourced platforms: content analysis of social media photographs. Open Geosci. 11, 558–571. https://doi.org/10.1515/geo-2019-0046.

Cayla, N., 2014. Volcanic geotourism in France: new perspectives for geotourism development. In: Erfurt-Cooper, P. (Ed.), Volcanic Tourist Destinations, Geoheritage, Geoparks and Geotourism. Springer, Berlin Heidelberg, Berlin, Heidelberg, pp. 131–138. https://doi.org/10.1007/978-3-642-16191-9_9.

Chai-allah, A., Fox, N., Brunschwig, G., Bimonte, S., Joly, F., 2023. A trail-based approach using crowdsourced data to assess recreationists' preferences for landscape. Landsc. Urban Plan. 233, 104700. https://doi.org/10.1016/j.landurbplan.2023.104700.

Cheng, X., Van Damme, S., Li, L., Uyttenhove, P., 2019. Evaluation of cultural ecosystem services: a review of methods. Ecosyst. Serv. 37, 100925. https://doi.org/10.1016/j.ecoser.2019.100925.

Chhetri, P., 2015. A GIS methodology for modelling hiking experiences in the Grampians National Park, Australia. Tour. Geogr. 17, 795–814. https://doi.org/10.1080/14616688.2015.1083609.

Corine Land Cover, 2018. https://land.copernicus.eu/pan-european/corine-land-cover/clc2018.

Costanza, R., 2008. Ecosystem services: multiple classification systems are needed. Biol. Conserv. 141, 350–352. https://doi.org/10.1016/j.biocon.2007.12.020.

Devlin, J., Chang, M.-W., Lee, K., Toutanova, K., 2018. {BERT:} Pre-training of Deep Bidirectional Transformers for Language Understanding. CoRR abs/1810.0.

Edwards, T., Jones, C.B., Corcoran, P., 2022. Identifying wildlife observations on Twitter. Ecol. Inform. 67, 101500. https://doi.org/10.1016/j.ecoinf.2021.101500.

Egarter Vigl, L., Marsoner, T., Giombini, V., Pecher, C., Simion, H., Stemle, E., Tasser, E., Depellegrin, D., 2021. Harnessing artificial intelligence technology and social media data to support cultural ecosystem service assessments. People Nat. 3, 673–685. https://doi.org/10.1002/pan3.10199.

Feinerer, I., Hornik, K., 2018. *tm: Text mining package*. R package version 0.7–6. Retrieved from. https://CRAN.R-proje. ct.org/package=tm.

Fox, N., Graham, L.J., Eigenbrod, F., Bullock, J.M., Parks, K.E., 2021a. Enriching social media data allows a more robust representation of cultural ecosystem services. Ecosyst. Serv. 50, 101328. https://doi.org/10.1016/j.ecoser.2021.101328.

Fox, N., Graham, L.J., Eigenbrod, F., Bullock, J.M., Parks, K.E., 2021b. Reddit: a novel data source for cultural ecosystem service studies. Ecosyst. Serv. 50, 101331. https://doi.org/10.1016/j.ecoser.2021.101331.

Ghermandi, A., Sinclair, M., 2019. Passive crowdsourcing of social media in environmental research: a systematic map. Glob. Environ. Chang. 55, 36–47. https://doi.org/10.1016/j.gloenvcha.2019.02.003.

Ghermandi, A., Langemeyer, J., Van Berkel, D., Calcagni, F., Depietri, Y., Egarter Vigl, L., Fox, N., Havinga, I., Jäger, H., Kaiser, N., Karasov, O., McPhearson, T., Podschun, S., Ruiz-Frau, A., Sinclair, M., Venohr, M., Wood, S.A., 2023. Social media data for environmental sustainability: a critical review of opportunities, threats, and ethical use. One Earth 6, 236–250. https://doi.org/10.1016/j.oneear.2023.02.008.

Gobster, P.H., Westphal, L.M., 2004. The human dimensions of urban greenways: planning for recreation and related experiences. Landsc. Urban Plan. 68, 147–165. https://doi.org/10.1016/S0169-2046(03)00162-2.

Gugulica, M., Burghardt, D., 2023. Mapping indicators of cultural ecosystem services use in urban green spaces based on text classification of geosocial media data. Ecosyst. Serv. 60, 101508. https://doi.org/10.1016/j.ecoser.2022.101508.

Günther, F., Rinaldi, L., 2022. Language statistics as a window into mental representations. Sci. Rep. 12, 8043. https://doi.org/10.1038/s41598-022-12027-5.

Günther, F., Dudschig, C., Kaup, B., 2015. LSAfun - an R package for computations based on latent semantic analysis. Behav. Res. Ther. 47, 930–944. https://doi.org/10.3758/s13428-014-0529-0.

Günther, F., Rinaldi, L., Marelli, M., 2019. Vector-space models of semantic representation from a cognitive perspective: a discussion of common misconceptions. Perspect. Psychol. Sci. 14, 1006–1033. https://doi.org/10.1177/1745691619861372.

Haines-Young, R., Potschin, M.B., 2018. Common International Classification of Ecosystem Services (CICES) V5.1: Guidance on the Application of the Revised Structure. Nottingham, United Kingdom.

Hansmann, R., Hug, S.-M., Seeland, K., 2007. Restoration and stress relief through physical activities in forests and parks. Urban For. Urban Green. 6, 213–225. https://doi.org/10.1016/j.ufug.2007.08.004.

Hartmann, M.C., Schott, M., Dsouza, A., Metz, Y., Volpi, M., Purves, R.S., 2022. A text and image analysis workflow using citizen science data to extract relevant social media records: combining red kite observations from Flickr, eBird and iNaturalist. Ecol. Inform. 71, 101782. https://doi.org/10.1016/j.ecoinf.2022.101782.

Hausmann, A., Toivonen, T., Fink, C., Heikinheimo, V., Kulkarni, R., Tenkanen, H., Di Minin, E., 2020. Understanding sentiment of national park visitors from social media data. People Nat. 2, 750–760. https://doi.org/10.1002/pan3.10130.

Havinga, I., Bogaart, P.W., Hein, L., Tuia, D., 2020. Defining and spatially modelling cultural ecosystem services using crowdsourced data. Ecosyst. Serv 43. https://doi.org/10.1016/j.ecoser.2020.101091. Article 101091.

Heikinheimo, V., Di Minin, E., Tenkanen, H., Hausmann, A., Erkkonen, J., Toivonen, T., 2017. User-generated geographic information for visitor monitoring in a national park: A comparison of social media data and visitor survey. ISPRS Int. J. Geo Informa. 6, 85. https://doi.org/10.3390/ijgi6030085.

Hermes, J., Van Berkel, D., Burkhard, B., Plieninger, T., Fagerholm, N., von Haaren, C., Albert, C., 2018. Assessment and valuation of recreational ecosystem services of landscapes. Ecosyst. Serv. 31, 289–295. https://doi.org/10.1016/j.ecoser.2018.04.011.

Hirschberg, J., Manning, C.D., 2015. Advances in natural language processing. Science 349, 261–266. https://doi.org/10.1126/science.aaa8685.

Howley, P., 2011. Landscape aesthetics: assessing the general publics' preferences towards rural landscapes. Ecol. Econ. 72, 161–169. https://doi.org/10.1016/j.ecolecon.2011.09.026.

Huai, S., Van de Voorde, T., 2022. Which environmental features contribute to positive and negative perceptions of urban parks? A cross-cultural comparison using online reviews and Natural Language Processing methods. Landsc. Urban Plan. 218, 104307. https://doi.org/10.1016/j.landurbplan.2021.104307.

Hunter, S.B., Mathews, F., Weeds, J., 2023. Using hierarchical text classification to investigate the utility of machine learning in automating online analyses of wildlife exploitation. Ecol. Inform. 75, 102076. https://doi.org/10.1016/j.ecoinf.2023.102076.

Hussain, R.I., Walcher, R., Eder, R., Allex, B., Wallner, P., Hutter, H.-P., Bauer, N., Arnberger, A., Zaller, J.G., Frank, T., 2019. Management of mountainous meadows associated with biodiversity attributes, perceived health benefits and cultural ecosystem services. Sci. Rep. 9, 14977. https://doi.org/10.1038/s41598-019-51571-5.

Kaiser, N.N., Ghermandi, A., Feld, C.K., Hershkovitz, Y., Palt, M., Stoll, S., 2021. Societal benefits of river restoration – implications from social media analysis. Ecosyst. Serv. 50, 101317. https://doi.org/10.1016/j.ecoser.2021.101317.

Kaltenborn, B.P., Bjerke, T., 2002. Associations between environmental value orientations and landscape preferences. Landsc. Urban Plan. 59, 1–11. https://doi.org/10.1016/S0169-2046(01)00243-2.

Lampinen, J., Tuomi, M., Fischer, L.K., Neuenkamp, L., Alday, J.G., Bucharova, A., Cancellieri, L., Casado-Arzuaga, I., Čeplová, N., Cerveró, L., Deák, B., Eriksson, O., Fellowes, M.D.E., de Manuel, B.F., Filibeck, G., González-Guzmán, A., Hinojosa, M. B., Kowarik, I., Lumbierres, B., Miguel, A., Pardo, R., Pons, X., Rodríguez-García, E., Schröder, R., Sperandii, M.G., Unterweger, P., Valkó, O., Vázquez, V., Klaus, V.H., 2021. Acceptance of near-natural greenspace management relates to ecological and socio-cultural assigned values among European urbanites. Basic Appl. Ecol. 50, 119–131. https://doi.org/10.1016/j.baae.2020.10.006.

Lee, H., Seo, B., Koellner, T., Lautenbach, S., 2019. Mapping cultural ecosystem services 2.0 – potential and shortcomings from unlabeled crowdsourced images. Ecol. Indic. 96, 505–515. https://doi.org/10.1016/j.ecolind.2018.08.035.

Levy, O., Goldberg, Y., Dagan, I., 2015. Improving distributional similarity with lessons learned from word embeddings. TACL 3, 211–225. https://doi.org/10.1162/tacl_a_00134.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G., Dean, J., 2013. Distributed representations of words and phrases and their compositionality. ArXiv Preprint.

Müllner, D., 2013. Fastcluster: fast hierarchical, agglomerative clustering routines for R and Python. J. Stat. Softw. 53, 1–18.

Naili, M., Chaibi, A.H., Ben Ghezala, H.H., 2017. Comparative study of word embedding methods in topic segmentation. Proc. Comput. Sci. 112, 340–349. https://doi.org/10.1016/j.procs.2017.08.009.

Norman, P., Pickering, C.M., 2019. Factors influencing park popularity for mountain bikers, walkers and runners as indicated by social media route data. J. Environ. Manag. 249, 109413. https://doi.org/10.1016/j.jenvman.2019.109413.

Oakes, M.P., Farrow, M., 2006. Use of the chi-squared test to examine vocabulary differences in English language corpora representing seven different countries. Literary Linguist. Comput. 22 (1), 85–99. https://doi.org/10.1093/llc/fql044.

Pickering, C., Walden-Schreiner, C., Barros, A., Rossi, S.D., 2020. Using social media images and text to examine how tourists view and value the highest mountain in Australia. J. Outdoor Recreat. Tour. 29, 100252. https://doi.org/10.1016/j.jort.2019.100252.

Plieninger, T., Dijks, S., Oteros-Rozas, E., Bieling, C., 2013. Assessing, mapping, and quantifying cultural ecosystem services at community level. Land Use Policy 33, 118–129. https://doi.org/10.1016/j.landusepol.2012.12.013.

Purves, R.S., Koblet, O., Adams, B., 2022. Analysing environmental narratives computationally. In: Purves, R.S., Koblet, O., Adams, B. (Eds.), Unlocking Environmental Narratives: Towards Understanding Human Environment Interactions through Computational Text Analysis. Ubiquity Press, pp. 43–84. https://doi.org/10.5334/bcs.c.

R Core Team, 2021. R: A language and environment for statistical computing. R Foundation for Statistical Computing.

Richins, H., Hull, J., 2016. Mountain Tourism: Experiences, Communities, Environments and Sustainable Futures. Cabi.

Romolini, M., Ryan, R.L., Simso, E.R., Strauss, E.G., 2019. Visitors' attachment to urban parks in Los Angeles, CA. Urban For. Urban Green. 41, 118–126. https://doi.org/10.1016/j.ufug.2019.03.015.

Sahlgren, M., 2008. The distributional hypothesis. Rivista di Linguista 20, 33–53.

Schirpke, U., Meisch, C., Marsoner, T., Tappeiner, U., 2018. Revealing spatial and temporal patterns of outdoor recreation in the European Alps and their surroundings. Ecosyst. Serv. 31, 336–350. https://doi.org/10.1016/j.ecoser.2017.11.017.

Schirpke, U., Tasser, E., Ebner, M., Tappeiner, U., 2021. What can geotagged photographs tell us about cultural ecosystem services of lakes? Ecosyst. Serv. 51, 101354. https://doi.org/10.1016/j.ecoser.2021.101354.

Schirpke, U., Ghermandi, A., Sinclair, M., Van Berkel, D., Fox, N., Vargas, L., Willemen, L., 2023. Emerging technologies for assessing ecosystem services: a synthesis of opportunities and challenges. Ecosyst. Serv. 63, 101558. https://doi.org/10.1016/j.ecoser.2023.101558.

Sebei, H., Hadj Taieb, M.A., Ben Aouicha, M., 2018. Review of social media analytics process and big data pipeline. Soc. Netw. Anal. Min. 8, 30. https://doi.org/10.1007/s13278-018-0507-0.

Silge, J., Robinson, D., 2017. Text Mining with R: A Tidy Approach. O'Reilly Media. https://www.tidytextmining.com/index.html.

Soga, M., Gaston, K.J., 2020. The ecology of human–nature interactions. Proc. R. Soc. B 287, 20191882. https://doi.org/10.1098/rspb.2019.1882.

Tibshirani, R., Walther, G., Hastie, T., 2001. Estimating the number of clusters in a data set via the gap statistic. J. R. Stat. Soc. Ser. B Stat Methodol. 63, 411–423. https://doi.org/10.1111/1467-9868.00293.

Van Berkel, D.B., Tabrizian, P., Dorning, M.A., Smart, L., Newcomb, D., Mehaffey, M., Neale, A., Meentemeyer, R.K., 2018. Quantifying the visual-sensory landscape qualities that contribute to cultural ecosystem services using social media and LiDAR. Ecosyst. Serv. 31, 326–335. https://doi.org/10.1016/j.ecoser.2018.03.022.

Venter, Z.S., Gundersen, V., Scott, S.L., Barton, D.N., 2023. Bias and precision of crowdsourced recreational activity data from Strava. Landsc. Urban Plan. 232, 104686. https://doi.org/10.1016/j.landurbplan.2023.104686.

Wan, C., Shen, G.Q., Choi, S., 2021. Eliciting users' preferences and values in urban parks: evidence from analyzing social media data from Hong Kong. Urban For. Urban Green. 62, 127172. https://doi.org/10.1016/j.ufug.2021.127172.

Wartmann, F.M., Purves, R.S., 2018. Investigating sense of place as a cultural ecosystem service in different landscapes through the lens of language. Landsc. Urban Plan. 175, 169–183. https://doi.org/10.1016/j.landurbplan.2018.03.021.

Wartmann, F.M., Koblet, O., Purves, R.S., 2021. Assessing experienced tranquillity through natural language processing and landscape ecology measures. Landsc. Ecol. 36, 2347–2365. https://doi.org/10.1007/s10980-020-01181-8.

Wilcer, S.R., Larson, L.R., Hallo, J.C., Baldwin, E., 2019. Exploring the Diverse Motivations of Day Hikers: Implications for Hike Marketing and Management. JPRA. https://doi.org/10.18666/JPRA-2019-9176.

Wolf, I.D., Wohlfart, T., 2014. Walking, hiking and running in parks: a multidisciplinary assessment of health and well-being benefits. Landsc. Urban Plan. 130, 89–103. https://doi.org/10.1016/j.landurbplan.2014.06.006.